R E P O R T R E S U M E S

ED 003 062 24

ADMINISTERING A LISTENING COMPREHENSION TEST THROUGH USE OF TEACHER-READERS, SOUND FILM, AND TAPE RECORDINGS.

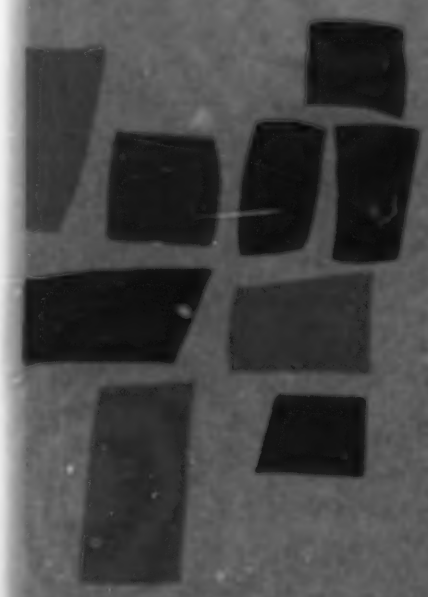
BY- STODOLA, QUENTIN AND OTHERS
NORTH DAKOTA STATE UNIV., FARGO

REPORT NUMBER CRP-1266 PUB DATE 62
NORTH DAKOTA UNIV., GRAND FORKS
REPORT NUMBER OE-10-020
EDRS PRICE MF-$0.09 HC-$1.68 42P.

DESCRIPTORS- LISTENING, *LISTENING SKILLS, *READING ABILITY, READING LEVELS, TEACHER QUALIFICATIONS, PUBLIC SCHOOLS, *AUDITORY VISUAL TESTS, FILMS, TAPE RECORDINGS, *EFFECTIVE TEACHING, TEACHING QUALITY, *COMPREHENSION DEVELOPMENT, FARGO, NORTH DAKOTA, GRAND FORKS, LISTENING COMPREHENSION TESTS OF THE SEQUENTIAL TESTS OF EDUCATIONAL PROGRESS

AN EXPERIMENT WAS CONDUCTED IN A TYPICAL SCHOOL SITUATION TO DETERMINE IF VARIATIONS AMONG TEACHER PRESENTATIONS OF LISTENING TESTS (READING ABILITIES) HAVE DIFFERENT EFFECTS ON THE RESULTS OF THESE TESTS. LISTENING TESTS WERE ADMINISTERED TO PUPILS OF THE PUBLIC SCHOOLS IN FARGO, NORTH DAKOTA, AT THE HIGH SCHOOL, JUNIOR HIGH SCHOOL, AND ELEMENTARY SCHOOL LEVELS. AT EACH GRADE LEVEL, A LISTENING COMPREHENSION TEST WAS GIVEN BY FOUR DIFFERENT METHODS TO GROUPS OF FIVE CLASS UNITS TO WHICH PUPILS AND TEACHERS HAD BEEN RANDOMLY ASSIGNED. THE METHODS WERE--(1) THE REGULAR CLASSROOM TEACHER READ THE TEST FOLLOWING THE TEST PUBLISHER'S DIRECTIONS, (2) THE CLASSROOM TEACHER READ THE TEST AFTER RECEIVING SPECIAL TRAINING FOR 2 HOURS, (3) THE CLASSROOM TEACHER GAVE THE TEST USING A PREPARED PROFESSIONAL MOTION PICTURE WITH SOUND, AND (4) THE CLASSROOM TEACHER ADMINISTERED THE TEST USING A PREPARED RECORDING, DUPLICATING THE AUDIO PORTION OF THE MOVIE. MEAN SCORES WERE DETERMINED FOR EACH CLASS UNIT AND SUBJECTED TO ANALYSIS OF VARIANCE. FOR ALL LEVELS AND UNITS, THERE WAS LITTLE EVIDENCE OF PRESENTATION DIFFERENCES CAUSING UNDUE VARIATION IN TEST SCORES. RELATIVE TO METHODS, ONLY AT THE HIGH SCHOOL LEVEL WAS THERE A MAJOR DIFFERENCE IN SCORES. HERE THE MOVIE GROUP SCORED SIGNIFICANTLY HIGHER THAN OTHER GROUPS, INDICATING GREATER PUPIL COMPREHENSION. (JC)

Cooperative Research Project No. OE-10-020



Administering A Listening Comprehension Test Through Use of Teacher-Readers, Sound Film and Tape Recordings

QUENTIN STODOLA • DONALD F. SCHWARTZ • RALPH H. KOLSTOE

1962 • NORTH DAKOTA STATE UNIVERSITY • FARGO

Table of Contents

	Page
THE MEASUREMENT OF LISTENING SKILLS	3-4
The Problem	
Previous Research	
THE RESEARCH PLAN	5-7
The Testing Instrument	
Methods of Test Administration	
The Sample Population	
Experimental Design	
MAKING THE MOTION PICTURE FILMS AND TAPE RECORDINGS	8-9
TESTING THE STUDENTS	10-12
High School Testing	
Junior High School Testing	
Elementary School Testing	
Equipment Used	
RESULTS	13-20
High School	
Junior High School	
Elementary School	
DISCUSSION OF RESULTS	21-24
Suggestions for Further Research	
Advantages of the Four Different Methods of Administering the Test	
SUMMARY	25-26
REFERENCES	26
APPENDIX	27-36
LIST OF TABLES AND GRAPHS	37

U.S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Administering A Listening Comprehension Test Through Use of Teacher-Readers, Sound Film and Tape Recordings

Prepared by:
Quentin Stodola
Donald F. Schwartz
NORTH DAKOTA STATE UNIVERSITY

Ralph H. Kolstoe
UNIVERSITY OF NORTH DAKOTA

1962

ED003062

The research reported herein was supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

Acknowledgements

Appreciation is expressed to the following persons on the staff of North Dakota State University who contributed valuable service to this research project: Donald A. Berreth, Clayton N. Haugse, John Hove, Paul A. Melchert, LaVerne W. Nelson, Donald M. Peterson, and Beth Rochefort.

Special thanks is offered to Superintendent Hamilton G. Vasey of the Fargo Public Schools, his administrative staff, the teachers and pupils. Without their splendid cooperation this research would not have been possible.

The Measurement of Listening Skills

THE PROBLEM

With the increased interest in improving listening habits, there is a need to be sure that scores on presently available listening tests are accurate indicators of pupil listening ability. Teachers want to train pupils how to listen effectively, but how can teachers plan lessons to improve listening if they are not able to measure present skill in listening with confidence? Furthermore, how can research workers measure comparative success of various ways of teaching listening if they are not able to trust test results?

In the listening tests presently available, the teacher reads selections aloud to the pupils and then asks questions on what has been read. In this kind of test administration, what is the danger of excessive variation in presentation causing a variation in test scores quite apart from the listening ability measured? In other words, would pupils get higher scores if they were tested by a good reader rather than by a poor one?

It seems obvious that an extremely poor reading, such as one spoken so softly that it could not be heard, would greatly lower student scores on a listening test. However, the question is, given normal variation in teacher reading ability, would there be enough difference between "good" and "poor" readers so as to result in significant differences in test scores.

The object of the experiment, then, was to find out whether or not in a typical school situation there is enough variation in reading abilities among teachers so as to have a significant effect on test results. Furthermore, it was hoped to try out some methods other than the usual teacher-reader type administration to determine whether or not undue variation in scores could be decreased.

PREVIOUS RESEARCH

Only during the past few years has there been specific research on the problem of the effect that differences among readers might have on listening test scores. Research has been generally limited to small groups and the results have been conflicting.

Berghley (1952) reports that vocally skilled speakers help listeners achieve higher scores on a listening test. On the other hand, Haberland (1958) found that the administration of a listening test by persons trained in speech did not give results significantly different from those obtained when the test was administered by persons untrained in speech.

In 1956, an experiment was conducted at Pennsylvania State University in which a listening test was administered to twelve different groups of students (Coffman, William E. and Quentin Stodola). The method of administering the test was varied from class to class, some classes having the reader of the script in the room, some viewing the reader on closed circuit TV, and some receiving only the sound from the TV set. Significant differences in mean test scores were found among the classes with different readers. But the use of different media of presentation (live, TV, sound) did not appear to affect the level of test scores.

The limited research on the effects of the differences in readers and the conflicting results indicate the need for the present experiment.

The Research Plan

THE TESTING INSTRUMENT

The Listening Comprehension Tests of the Sequential Tests of Educational Progress (Educational Testing Service, Princeton, New Jersey) were used in this experiment.* These tests are available in four levels designed for grades four through fourteen. For this experiment, the tests were used at three levels, grades four through twelve.

Administration of an entire test takes seventy minutes with two thirty-five minute sections, each section paralleling the other in content and difficulty. For this experiment, the first thirty-five minute sections of the tests were used at all three levels.

In these tests, the student answers questions about a series of passages read to him. The questions for each passage are read immediately after the passage. The answer options, not the entire questions, are available to the student in a test booklet during the test. The selections represent typical listening situations which the student may encounter in school. They include directions, explanations, stories, arguments, poems, etc. The questions require the student to select the important facts and think critically about them.

METHODS OF TEST ADMINISTRATION

How great are the effects on listening test scores caused by differences among readers? One way to approach this problem is to compare variations in test scores obtained from administrations under typical classroom conditions with teacher-readers, as contrasted with test administrations where differences among readers are presumably eliminated or at least minimized. Accordingly, it was decided to administer the test under four conditions, the first with teacher-readers in typical classroom situations. The other three methods were planned to help eliminate differences among readers as much as possible. It was thought that one of these three methods might provide more stable test results.

These methods of administration are described as follows:

1. REGULAR

The REGULAR classroom teacher reads the test following

*The two most commonly used tests of listening comprehension are *The Brown-Carlsen Listening Comprehension Test* (World Book Company) and *The Listening Comprehension Test of the Sequential Tests of Educational Progress* (Educational Testing Service). *The Brown-Carlsen Test* is appropriate for grades nine through thirteen only.

the directions which are published with the test. Directions require the teacher to practice reading the selections and questions before administering the test.

2. TRAINED
   The classroom teacher reads the test after receiving a special two hour training session including supervised practice in reading the test. Basically, the purpose of the training is to make sure that the teachers actually do follow publisher's directions for giving the test, which include studying the test manual and practicing the selections at the suggested rate. The training does not involve any attempt to change basic speech patterns.
3. MOVIE
   The classroom teacher administers the test using a prepared sound motion picture where the test is given by a professional reader. The film is shown in a number of different classrooms using different projectors consistent with normal variations.
4. AUDIO
   The classroom teacher administers the test using a prepared tape recording which duplicates the audio portion (sound track) of the motion picture.

THE SAMPLE POPULATION

Substantially all the pupils in grades four through twelve of the Fargo Public Schools in Fargo, North Dakota, served as the population. Excellent cooperation in the project was provided by Superintendent Hamilton G. Vasey, his administrative staff, and the Fargo teachers.

Generally speaking, the teachers and the pupils of the Fargo Public Schools seem fairly typical of those found in many other school systems, although standardized test scores (Iowa Tests of Basic Skills and Iowa Tests of Educational Development) suggest above average achievement in school subjects for Fargo.

In explaining to teachers and pupils the reasons for giving the test, emphasis was placed on practical use of test results for instructional purposes, as well as on the underlying experimental nature of the test administrations. Test scores were provided for the teachers within three weeks after the test administration, and principals and curriculum consultants encouraged use of the scores by the teachers for better understanding of the specific listening problems of individual students and also as a basis for providing classroom instruction to improve listening.

Results of the experiment were reported to the teachers and community through a half-hour television program and through an article in the NORTH DAKOTA TEACHER, the magazine of the state education association.

EXPERIMENTAL DESIGN

The teacher and pupils comprising each class unit were selected to participate in the experiment on a random basis using a table of random numbers within the grade levels. Thus, completely new class units were created especially for this experiment.

The following table illustrates the distribution of the experimental group:

Number of Classes (Approximately 25 students per class)

Grade		REGULAR (Teacher reads following published directions)	TRAINED (Teacher reads after training session)	MOVIE	AUDIO (Tape Recording)
High	12	5	5	5	5
School	11	5	5	5	5
	10	5	5	5	5
Junior	9	5	5	5	5
High	8	5	5	5	5
	7	5	5	5	5
Elem.	6	5	5	5	5
	5	5	5	5	5
	4	5	5	5	5

In each case, the class unit was used as the analysis unit. Thus, a mean score was determined for each class unit in each grade level-treatment combination. This mean score was used as the score for analysis purposes and all statistical treatment consisted of analysis of differences among these scores.

Variance among the scores of the class units for each level (High School, Junior High, Elementary) was analyzed to determine whether there was undue variation which might be ascribed to the method of administration. An analysis was also made to determine whether there were significant differences among the mean scores according to method.

Generally speaking, the approach may be described as analyses of variance consisting of a combination of randomized replications and a randomized block design with the grade levels constituting the blocks and the groups within grade levels and within treatments (across grade levels) constituting the replications.

Making The Motion Picture Films and Tape Recordings

The films and tapes were produced by Bill Snyder Films, a local professional company which has complete facilities for 16 mm. film production. Films produced by the company have been award winners in international film competition.

Special problems were encountered in production of the film. The question arose as to whether or not the film should simply attempt to parallel the kind of presentation that might be given by a typical classroom teacher or whether the film version of the test should present a professionally trained speaker who presumably would take full advantage of his reading skills to communicate effectively. It was decided that the second kind of presentation using the professional speaker would make the more desirable listening test. Accordingly, Glenn Hanson, a highly competent television news reporter was selected to read the test for filming.

The narrator was given the following instructions:

1. When reading the general instructions to the test at the beginning, be informal, pleasant, and relaxed in manner so as to help put the students at ease. Try to create the feeling that the film will provide an enjoyable listening experience. Think of your role as one who will be presenting interesting selections rather than as one giving a test.
2. Use facial expressions and hand gestures as seem natural to you, consistent with the mood of the reading selection.
3. The mood of the different selections should be varied through reading rate, vocal intonation, gesture, and bodily posture; however, you should only suggest the role rather than act it out. Avoid giving a highly dramatic presentation.
4. Occasionally use some bodily movement to suggest transition from one role to another. (For example, speaker in one selection to questioner on that selection, to speaker in the next selection, and so forth.)

A table and chair were used as props. Generally speaking, the more serious and formal selections were given with the narrator standing, those which were professorial or academic in tone with the narrator seated behind the table, and those which were informal with the narrator either seated beside the table or on the edge of the table.

With the exception of certain poetic selections which were read from a book held by the narrator, the selections were read from a teleprompter to increase eye contact and effect intimacy. The questions were read from the test booklet as would ordinarily be done by the classroom teacher giving the test.

In the film version of the test, the students were not provided with booklets giving the answer options. Instead the options appeared on the motion picture screen as the narrator read them, and remained there for time periods varying from ten to fifteen seconds, according to item difficulty as determined by analysis supplied by the publisher and according to the experimenters' subjective judgment of the difficulty of the item. The time allowances were pretested on small groups and appeared satisfactory.

The lengths of the films at the three levels are as follows:

High School (10-12) — 44 minutes
Junior High (7 - 9) — 40 minutes
Elementary (4 - 6) — 39 minutes

The films were processed to provide an especially bright picture so that it was possible to allow sufficient room light for students to read the answer sheets.

In order to standardize the comparison between the tape recordings and films, the original film sound track was dubbed to magnetic tape. By so doing, the AUDIO portion of the test was an exact duplicate as to timing and vocal expression. The films were made by the double system filming method to obtain film prints which have the highest quality optical sound track and tape recordings of optimum fidelity.

Testing The Students

It is believed that the problems encountered in the test administration are typical of those which might be expected in many school systems. In most classes, the administrations went smoothly and standardized procedures were followed. Although discussion will center on the problems encountered, it should be kept in mind that these represent exceptions rather than the rule.

HIGH SCHOOL TESTING

At the high school level, it was necessary to assign more than eleven hundred students to sixty experimental groups on a random basis. The students were given their assignments in advance of the testing day and reported to their rooms with relatively little confusion. Unfortunately, because of a clerical error, one of the sixty groups did not report to the room assigned for testing at the designated time and it was thought best to eliminate them from the experiment. So only fifty-nine experimental groups were used at the high school level.

Information about the testing was obtained through listening on the intercom system, seating observers in some of the MOVIE and AUDIO rooms, and having each teacher fill out a short questionnaire to indicate unusual incidents.

The following instances of non-standardized procedure were noted through listening over the intercom. Two teachers didn't read the option letters to the various questions and one teacher didn't read the option statements at all, but depended on pupils to read the statements silently from their own booklets. This latter teacher repeated some of the questions and his reading was very dramatic. All three teachers previously noted were in the administration group designated as REGULAR where teachers administer the test without special help following the publisher's directions.

The following are some statements from the teacher questionnaire indicating unusual incidents:

"We could hear the tape recorder from the adjoining room."

"We heard dishes rattling from the kitchen."

"The projector did not work at the start and had to be stopped and readjusted."

"The tape recorder had a squeal in parts of the test and the recorder in the next room also gave some interference."

"The speed of the projector was too fast so that the film came loose and had be readjusted."

"One girl had a coughing spell. I gave her a drink of water and a cough drop."

Observation was made of three film (MOVIE) and three tape recording (AUDIO) administrations of the test. The observers were a college English teacher and two Psychology teachers. Generally speaking, they thought that the administrations went very smoothly and the morale of the students was excellent. They did, however, note a few minor problems such as broken pencils and the need to change the angle of a motion picture screen slightly.

JUNIOR HIGH SCHOOL TESTING

Observation of testing was made through the intercom system and use of teacher questionnaires.

Again most administrations went smoothly. However, a technical problem developed in connection with one of the motion picture administrations. Because of the stiffness of the new film, the film in one room kept jumping off the sprockets of the projector causing the machine to be stopped over thirty times during the test. These stops came in the middle of selections and questions. It was tentatively decided not to include results of this administration in the experiment because the administration was so atypical. Surprisingly enough, the statistical analysis revealed no important differences between this group and the other four groups who took the MOVIE administration at the same grade level. Therefore, it was decided to use the results in the overall statistical analysis. (The mean score and variance for this class is given in Table B of the Appendix. The class is the first one listed for the seventh grade level under the MOVIE method of administration.)

The experimenters can only surmise why the pupils in this particular group were able to do reasonably well in answering the questions despite the handicap of repeated interruptions during the test. Perhaps pupils who are highly motivated and who are trying to do well on a test can often screen out distractions and concentrate on the task at hand. In this case, the selections were not extremely difficult nor were most of the questions very challenging. (Table D of the Appendix indicates the average difficulty of the questions at the various grade levels.) The interruptions in this instance were handled by the students without undue difficulty as far as could be ascertained.

Here are some instances of non-standardized procedures indicated by the the teachers who completed questionnaires:

"The noise of the tape recorder next door was disturbing to my pupils."

"Some of the girls thought the selection 'Shadow Box Eggs'

was humorous and therefore distracting, possibly because a man was reading it."

"There was an announcement over the public address system during the reading ~~of the~~ first selection."

"The tape recorder slowed down for awhile; the machine stopped and then started and the speed returned to normal."

However, most of the teachers commented that their administrations went smoothly.

THE ELEMENTARY SCHOOL TESTING

No questionnaires were given the teachers at the elementary school level since each school was tested separately and it was possible to talk to the relatively small number of teachers involved in each school on an individual basis. Only a few deviations from standardized administration were mentioned. One of the teachers had to stop the test for five minutes because a workman was making considerable noise with a power drill in the basement. In another case the film jumped off the sprockets of the projector and there was some delay in starting the film again.

Over the intercom system the investigators noted several instances of teachers reading the selections more slowly than suggested by the test publisher. At the fourth grade level, three REGULAR teachers and one TRAINED teacher read slowly and at the fifth grade level one TRAINED teacher was noted reading the test at a slow pace. One TRAINED teacher at the fifth grade level did not know that she needed test booklets with answer options for her pupils and had to be so informed just before she began reading the selections to her pupils. In reading the directions, one of the fourth grade REGULAR teachers kept repeating parts of the directions several times and before each selection she exhorted the students to listen carefully, even though this is not part of the published directions.

Generally speaking, there seemed to be a tendency for the teachers in the lower elementary grades, particularly the fourth grade, to read the selections more slowly than suggested by the test publisher. This slow reading is evidently consistent with the slow and precise speech that many teachers use in talking to their classes at the lower grade levels.

EQUIPMENT USED

The types of movie projectors and tape recorders used are typical of those available in many schools. The projectors used were all Bell & Howell Standard Specialist. A list of the tape recorders used is given in Table J of the Appendix.

Results

The results will be discussed separately for high school, junior high school, and elementary school. A different level of the test was used for each of these groups.

A detailed summary of test scores as well as other data is given in the Appendix. However, the main findings are listed in the following sections.

HIGH SCHOOL

The summary of the results of testing at the high school are given in Table I.

Table I

Means and Variances of Listening Test Scores by Method of Presentation for Grades 12, 11, and 10

GRADE		REGULAR	TRAINED	MOVIE	AUDIO	TOTALS
12	$\bar{X}$ (mean score)	23.9	24.1	26.9	24.9	25.0
	s^2 (variance)	.81	4.68	1.78	1.84	
	n	5	5	5	5	
11	$\bar{X}$	24.4	23.0	24.4	24.4	24.0
	s^2	1.85	1.42	.91	1.28	
	n	5	5	5	5	
10	$\bar{X}$	21.7	23.0	24.4	22.6	22.9
	s^2	.57	3.44	.92	.60	
	n	5	5	4	5	
TOTALS	$\bar{X}$	23.3	23.3	25.2	23.9	
	n	15	15	14	15	

The data in Table I were analyzed to determine (1) if the observed differences in variances were greater than that expected by sampling differences and (2) if the differences in means were beyond chance expectations.

Bartlett's Test was used to test for homogeneity of variance. The obtained Chi square (9.89 with 11 degrees of freedom) does not warrant rejection of the null hypothesis with respect to differences in variances. Whatever the method of administration, the results were more or less equally constant within methods.

It should be noted, from Table I, that two of the cell variances are considerably larger than the other ten (also see Table

E of the Appendix). These two extremes occurred with the TRAINED teachers at the tenth grade level and at the twelfth grade level. Further inspection indicates that the increased variance in each case is produced by one class unit (see Table F of the Appendix). In the case of the tenth grade, the score for one class unit is considerably higher than for the other four. At the twelfth grade, the score for one class unit is considerably lower than for the others.

In a sense then, after examining carefully the fifty-nine classes* in the twelve groups, only two scores were identified as varying markedly from the others in their group, possibly because of the method of administration. Since there were a relatively large number of classes, the two "unusual" classes may be explained on the basis of chance factors quite apart from the way the tests were administered. It may be simply a matter of chance in the sense that an inordinately large number of poor listeners were assigned to the low group in the twelfth grade; and conversely by chance a large number of good listeners may have been assigned to the class which did better than expected at the tenth grade. Thus, summing up the evidence, there does not seem to be much reason to believe that the REGULAR teacher-reader method of administration leads to any more variability in test scores than do the other three methods.

The second question—possible variation in means of the subgroups—was tested by use of a four by three analysis of variance. This question relates to whether or not any one of the four methods resulted in generally higher test scores than did the others. The results of this analysis are shown in Table II.

Table II

Grades 10-12

Analysis of Variance by Methods of Administration and Grade Levels

Source of Variation	df	SS	MS	F
Methods of Administration	3	35.72	11.91	6.96**
Grade Levels	2	41.40	20.70	12.11**
Methods X Grade Levels	6	18.68	3.11	1.82
Within Cells	47	80.42	1.71	

**Significant at .01 level

As indicated in Table II, the differences among methods of test administration were significant beyond the .01 level. As might be expected, highly significant differences among grade levels were also obtained. No interaction effects were noted between methods of administration and grade level.

*It may be recalled that one of the sixty classes originally planned to participate was eliminated because of a mix-up in scheduling.

To determine which methods of test administration were significantly different from the others, the Duncan's Multiple Range Test (Edwards, 1960) was applied to the mean scores of the different methods of administration. Results of the tests for grades ten through twelve are given in the Appendix as part of Table G. As indicated in Table G, the mean score for the MOVIE groups is significantly higher than for the other three methods at the .05 level and there are no significant differences among the mean scores of the other three methods. It would appear then that comprehension of the MOVIE version of the test is better than other methods.

The fact that students made higher scores with the MOVIE method should not be interpreted to mean that the MOVIE is a better testing instrument for measuring listening ability. From the point of view of statistical reliability, the less difficult test is not necessarily the best. The whole matter of the relative advantages and disadvantages of the four methods of administration will be discussed in a later section.

JUNIOR HIGH SCHOOL

The summary of the results of testing at the junior high schools are given in Table III.

Table III

Means and Variances of Listening Test Scores By Method of Presentation for Grades 9, 8, and 7

GRADE		REGULAR	TRAINED	MOVIE	AUDIO	TOTALS
9	$\overline{X}$ (mean score)	31.2	31.0	31.2	30.3	30.9
	s^2 (variance)	2.23	.87	.94	.70	
	n	5	5	5	5	
8	$\overline{X}$	30.0	29.7	30.4	29.5	29.9
	s^2	1.32	1.64	2.09	.51	
	n	5	5	5	5	
7	$\overline{X}$	27.4	27.6	27.0	27.8	27.5
	s^2	1.05	1.05	.97	1.91	
	n	5	5	5	5	
TOTALS	$\overline{X}$	29.5	29.4	29.5	29.2	
	n	15	15	15	15	

The data in Table III were analyzed in a manner similar to the high school group to determine (1) if the observed differences in variance were greater than that expected by sampling differences and (2) if the differences in means were beyond chance expectations.

Applying Bartlett's Test to the data, no significant differences among the variances of the class units within groups were observed. The obtained Chi Square (4.41 with 11 degrees of freedom) does not warrant rejection of the null hypothesis with respect to differences in variance.

Using the same procedure as with the high school, possible variation in means of the subgroups was tested by use of a four by three analysis of variance. The results of this analysis are shown in Table IV.

Table IV
Grades 7-9
Analysis of Variance by Methods of Administration and Grade Levels

Source of Variation	df	SS	MS	F
Methods of Administration	3	1.01	.37	.27
Grade Levels	2	127.94	63.97	50.27**
Methods X Grade Levels	6	5.23	.87	.69
Within Cells	48	61.07	1.27	

**Significant at .01 level

As indicated in Table IV, differences among methods of test administration were not significant. Nor were there any interaction effects between method of administration and grade level. As expected, differences in scores among grade levels were significant beyond the .01 level.

Thus, to sum up results for the junior high school, with this particular test, the method of administration does not appear to affect the variability of scores among class units nor the comprehension level.

ELEMENTARY SCHOOL

No elementary school in the system was large enough to permit replication of each treatment condition at each grade level. Nor (with the exception of one school at grade six) was it possible to administer each treatment condition at least once at any grade level.

This problem was obviously not encountered at the senior high school level where only one school was involved and all the students in the sample were available at one time and location. Nor was it a problem at the junior high school level where there were two schools, but these did not differ significantly in listening test scores.

In the analysis of the listening test scores in the elementary schools, possible differences in the methods of test administration and also difference in grade level are partially confounded by pos-

sible differences among schools. Results of the Iowa Tests of Basic Skills, which are administered every year in Fargo, provide evidence for differences among schools in the areas measured.

In order to correct for differences among schools at the elementary school level, the possibility of using an analysis of covariance procedure with grades on the Iowa Test of Basic Skills as the adjustment factor was considered. However, since achievement test scores do not seem to be highly correlated with the listening test, it is not certain that the Iowa Test of Basic Skills would have been an appropriate predictor variable (Baldauf, 1960). Also, there was the matter of the additional costs involved in assembling the achievement test scores and analyzing them. For these reasons, it was decided to use the analysis of variance approach as originally planned.

The summary of results of testing at the elementary schools is shown in Table V.

Table V

Means and Variances of Listening Test Scores By Method of Presentation for Grades 6, 5 and 4

GRADE		REGULAR	TRAINED	MOVIE	AUDIO	TOTALS
6	$\bar{X}$ (mean score)	34.5	34.3	33.4	32.4	33.7
	s^2 (variance)	1.31	1.42	1.79	1.49	
	n	5	5	5	5	
5	$\bar{X}$	31.5	30.6	32.3	32.0	31.6
	s^2	1.66	1.21	2.66	1.37	
	n	5	5	5	5	
4	$\bar{X}$	30.4	28.6	31.0	28.8	29.7
	s^2	1.46	.85	3.07	4.03	
	n	5	5	5	5	
TOTALS	$\bar{X}$	32.1	31.1	32.2	31.0	
	n	15	15	15	15	

Results of Bartlett's Test following the same method used at the senior and junior high school levels, did not indicate any significant variability among variances according to methods of administration by grade level. The obtained Chi Square was 4.49 with eleven degrees of freedom.

However, as indicated in Table E of the Appendix, when analysis of variance procedures were applied to groups of classes, using the scores for individual students as the unit of analysis, according to methods of administration by grade level, in six cases significant results were obtained at either the .05 or the .01 levels.

Since the results of Bartlett's Test applied to the whole group did not indicate heterogeneity of variance, the "significant"

results obtained with the six groups separately cannot be interpreted as indicating undue variability of scores within these groups. However, for reasons of interest, a further analysis of these six groups was made utilizing Duncan's Multiple Range Test for unequal n's (Kramer, 1956). The results of this analysis are given in Table F.

A study of Table F indicates that in only one of the six groups is there a "significant" difference in test scores within a school. In the other cases, one might suspect that the differences in test scores, rather than representing variability caused by the method of administration, is variability caused by differences in pupil abilities among schools. Thus the evidence of variability within methods is very limited.

Following the approach used at the senior and junior high school levels, analysis of variance of test scores was made according to method of administration by grade levels as indicated in Table VI.

Table VI

Grades 4-6

Analysis of Variance by Methods of Administration and Grade Levels

Source of Variation	df	SS	MS	F
Methods of Administration	3	17.05	5.68	3.05*
Grade Levels	2	155.72	77.86	41.86**
Methods X Grade Levels	6	26.14	4.36	2.34*
Within Cells	48	89.22	1.86	

*Significant at .05 level
**Significant at .01 level

Differences among methods were significant at the .05 level and, as expected, differences among grade levels were highly significant. An interaction effect significant at the .05 level cast some doubt on the direct interpretation of the main effects of grade level and of methods of administration.

Two different graphs of this significant interaction are presented in the Appendix on pages 35 and 36, (Graph 1 and Graph 2). Examination of Graph 1 indicates that even though the differences between grade levels at each method of administration are not proportional, the order of grade six is greater than grade five and is greater than grade four which is maintained throughout. Thus the main effects for grade level, as given in Table VI, are directly interpretable as far as indicating the order of simple effects across the different methods of test administration.

Graph 2 indicates that the differences in methods of test administration across grade level cannot be interpreted in any simple

manner. This graph shows not only how the rank of the various mean scores on the different methods of test administration change from grade level to grade level, but also how the comparative ranks of the schools involved vary on the Iowa Tests of Basic Skills. The rank of the mean scores for the different methods of test administration change from grade level to grade level without following any readily apparent pattern.

At the fourth grade level, although both the TRAINED teacher and MOVIE groups have the same rank on the Iowa Tests of Basic Skills, there is considerable spread between mean scores of the two groups with the MOVIE being higher. On the other hand, at the sixth grade level, the mean scores for the different methods of administration are consistent with scores made by the schools on the Iowa Tests of Basic Skills.

It would appear foolhardy to offer with much confidence any explanation for these interaction effects. Interpretation is complicated by the probability that differences among schools are affecting some of the mean scores obtained with the different methods of administration.

However, this hypothesis is tentatively suggested: At the fourth grade level, teachers speak to their pupils quite slowly and precisely. By the time the sixth grade level is reached the teachers speak to pupils in a manner more consistent with ordinary patterns of speech. In the testing situation, it may be that the TRAINED teachers, in attempting to follow directions, spoke at a more rapid pace than usual for these pupils. On the other hand, observation indicated that, despite test directions, those administering the test under REGULAR conditions tended to speak more slowly than the others. This may explain why the REGULAR teachers were comprehended better than the TRAINED teachers and the AUDIO at the fourth grade level.

Of course, in the fourth grade MOVIE where scores were relatively high, the presentation was consistent in rate with the AUDIO where scores were low. However the filmed visualization may have compensated for the rapid delivery.

The factors just mentioned may have not been operating at the sixth grade level where the students are accustomed to a more normal speed of delivery in speech. Again, to say the least, this hypothesis for some of the interaction effects between method and grade is offered in a tentative manner.

Actually the differences are not very great among the scores for the various methods of test administration by grade level. At

the fifth grade level, results of the Duncan Multiple Range Test reveal no significant differences at the .05 level of confidence. At the fourth and sixth grade levels, differences which are significant are indicated in Table H of the Appendix.

In-so-far as possible, scores for the various methods of administration were also compared within schools according to grade level by means of a "t" Test to identify significant differences among mean scores.

The results of these comparisons are indicated in Table I of the Appendix. As may be noted, in only three cases out of twelve were results significant at the .05 level obtained:

The combined average for two MOVIE classes was higher than for a TRAINED teacher class (fourth grade).

The combined average for two REGULAR classes was higher than for an AUDIO class (sixth grade).

The average for a TRAINED teacher class was higher than for two AUDIO classes combined (sixth grade).

Summing up the results for the elementary school level, there was no evidence of undue variation among scores for class units using the same methods of test administration. Nor was the variability of scores among class units using different methods of test administration especially high. Interaction effects between methods of administration and grade level were noted.

Discussion of Results

Under the conditions of this experiment, for all grade levels tested (four through twelve), significant results were not obtained to indicate excessive variability *within* methods of test administration. In analyzing the variance *among* methods of test administration, significant results were obtained at the high school and elementary school levels.

Scores for the MOVIE method of test administration (at the high school level) were significantly higher than for the other methods. Why was the MOVIE version of the test better understood at the high school level? The experiment itself provides no answer. Possibly at the upper grade levels where the test presumably requires high levels of reasoning, the advantage of listening to the professional reader of the MOVIE version, who gave an exceedingly meaningful presentation, helped raise student scores.

At the elementary school level, interaction effects between grade level and method, as well as possible differences among the schools involved, cloud the meaning of the results. The relationship between method of test administration and mean score is not clear.

SUGGESTIONS FOR FURTHER RESEARCH

The experimenters are keenly aware that the results are dependent on the measuring instrument used. As shown in Table D of the Appendix, the test questions were relatively easy for the students at all grade levels. For example, at the sixth grade level, the average score was 84.3 percent right.

The question arises: What if the test had been more difficult? Would differences among readers on a more difficult test have caused greater variation in test scores? Also would there have been greater differences in comprehension level according to the method of administration used? For example, would comprehension of the MOVIE version been proportionately better if the test had been harder?

An answer to these questions might be found by conducting another experiment paralleling the present one with the exception that the tests would be administered one level lower. This would mean that the test used at the high school level would be given to the junior high school students, and the test given to the junior high school would be given to the elementary school pupils. Of

course, there would be no test available for the senior high unless film and tape versions of the college level test were prepared, which could then be used at the high school level.

Because of the differences among schools, results at the elementary school level are confounded. It would be relatively simple to repeat the present experiment at the elementary school level within a single large elementary school, possibly breaking down some of the classes into groups of fifteen or twenty students in order to get the desired number of experimental groups. If this experiment were conducted, it would be possible to check variation in the comprehension level among the different methods of administration and also check extent of interaction between grade levels and methods of administration.

ADVANTAGES OF THE FOUR DIFFERENT METHODS OF ADMINISTERING THE TESTS

It was hoped that this experiment would provide clear-cut evidence as to which of the four methods was the best method of administering a listening test, particularly in regard to the problem of possible variability caused by differences in test readers. As it turned out, results indicate that with the particular test used, there was no undue deviation in scores which could be ascribed to any one of the four methods of giving the test. It did not appear that in a school situation similar to Fargo there need be special concern about the effect of reader differences causing undue variation in test scores.

If either the TRAINED teacher, the MOVIE, or AUDIO method of giving the test is better than the REGULAR method, it will have to be proven on some grounds other than that of deviation in test scores caused by reader differences.

For reasons of interest, a brief discussion follows concerning some of the advantages and disadvantages of the various methods of administering the test apart from the question of variability of scores.

1. REGULAR — Possibly the greatest advantage of this method is that it is more natural than either the MOVIE or AUDIO methods. The pupil listens to his own teacher. Neither the pattern of speech nor the mode of presentation is unnatural.

This method also has the advantage of involving the teacher directly in the testing process. Since the teacher has to practice reading the test and questions, she becomes more aware of exactly

what the test is attempting to measure and she can probably interpret test results better.

This method also permits a certain flexibility. That is, if there is some unusual happening, such as a student having a coughing spell, the teacher can stop, and if necessary, repeat part of a selection or question.

The obvious disadvantage to this method is that it does require considerable time and effort on the teacher's part to prepare for the test. Another disadvantage is that some teachers read the selections in a dull manner, causing pupils to become bored.

2. TRAINED — The purpose of this method was to make sure that the teachers actually followed the publisher's directions in getting ready for the test. Results of the experiment seemed to indicate that this special training was not necessary, that probably in fact the REGULAR teachers who prepared on their own were essentially as well prepared as the TRAINED teachers. Using this method would seem to constitute an extra and unnecessary burden for whoever might be in charge of the special training—probably the school principal or members of his staff. Thus there does not seem to be any special reason for using this method of administration.

3. MOVIE — The greatest advantage of the MOVIE method from the experimenters' point of view was the pleasure that the students seem to get from hearing the selections read by a professional reader. Very few of the teachers, indeed, approached the skill of the professional in the readings they gave.

Another advantage of the MOVIE version of the test is the fact that it is easier for the teacher since she does not have to practice the readings nor give them in the test situation.

With the MOVIE version there is no need to provide test booklets with answer options since these are shown on the screen.

An obvious disadvantage in giving the test by MOVIE is the problem of providing sufficient light for the pupils to mark the answer sheets and still have the room dark enough to view the picture. However, it is felt that this problem was not a serious one in the tests given in Fargo. Generally speaking, if most or all the lights in a room were turned off and the blinds at the front side of the room were pulled, a kind of half-light was provided that seemed optimum for the purpose.

Another disadvantage of the MOVIE method is the possibility of mechanical problems caused either by faulty projection equipment or by teachers inexperienced in projector operation.

The fact that the MOVIE version of the test seemed to be better comprehended and scores were higher is not cited as either an advantage or a disadvantage. It cannot be said that in this case a slightly easier test made a better measuring instrument.

The experimenters do not feel competent to attempt to compare costs of the MOVIE and REGULAR methods of administration. Obviously, it costs money to make and distribute film. It also costs money to print and distribute test booklets. Perhaps the extra time required by the teacher to practice and give the test should also be figured in a cost estimate comparing the REGULAR and MOVIE methods.

4. AUDIO — Of all the methods for giving the tests, using the tape recorder was probably easiest for the teacher. No practice in reading was needed beforehand, and tape recorders are simpler to set up and use than motion picture projectors. The fidelity of the sound is better than that of the MOVIE. The cost of making the tape is much less than the cost of a motion picture.

The absence of visual presentation seemed to make this method somewhat less stimulating and interesting to the pupils.

As indicated by the foregoing discussion, there are advantages and disadvantages to all of the methods of administration. No recommendation is made in favor of any one of these methods.

Summary

A major purpose of the experiment was to find out whether or not in a typical school situation there is enough variation in reading abilities among teachers giving a listening test so as to have a significant effect on test results. It was also hoped to find out if variability among scores for comparable class units could be reduced by using other methods of test administration.

Listening tests were administered to the Fargo Public School pupils at three different levels of difficulty: high school, junior high school, and elementary school. At each grade, the test was given by four different methods to groups of five class units to which pupils and teachers had been assigned on a random basis.

The four methods were as follows:

1. REGULAR
   The REGULAR classroom teacher read the test following the test publisher's directions.

2. TRAINED
   The classroom teacher read the test after receiving a special two hour training session including supervised practice in giving the test.

3. MOVIE
   The classroom teacher administered the test using a prepared sound motion picture where the test is given by a professional reader.

4. AUDIO
   The classroom teacher administered the test using a prepared tape recording duplicating the AUDIO portion of the the MOVIE.

Results of analysis of test scores are as follows:

1. At all three levels, there was no evidence of reader differences causing undue variation in test scores. In fact, the variability of mean scores by classes within all the methods of administration was not significantly high.

2. At the high school level, mean scores for the MOVIE groups were significantly higher than scores for the other groups (REGULAR, TRAINED, AUDIO) indicating greater comprehension by the MOVIE groups. At the junior high level, there was no significant difference among mean scores for the various methods of administration. Analysis of mean scores by methods at the elementary school level are not easily interpreted because of differences in student abilities according to school as well as the interaction effect noted in "3" following. However, it does not appear that the differences among the mean scores by method were especially high.

3. There was an interaction effect between grade levels and methods of administration at the elementary school level, but this interaction effect does not appear to follow a consistent or easily interpretable pattern.

A word of caution. Findings apply to the particular tests used in the Fargo school system. The tests were relatively easy. If more difficult tests had been used, results might have been different.

REFERENCES

Baldauf, Robert John. *A Study of a Measure of Listening Comprehension and It's Relation to the School Achievement of Fifth Grade Pupils*, Unpublished Doctoral Dissertation, University of Colorado, 1960

Burgbley, K. C. *Speech Monographs*, 19:249-258, 1952

Coffman, William E. and Quentin Shedola, "The Effect of Conditions of Administration on Listening Comprehension Test Scores," *Research Bulletin* 61-4, Educational Testing Service, Princeton, New Jersey, February, 1961

Edwards, Allen L. *Experimental Design in Psychological Research*, Rinehart & Company, Inc., New York, New York, 1960

Haberland, John A. "Speaker Effectiveness and The Brown-Carlsen Listening Test," *School and Society*, 86:186-188, April 26, 1958

Kramer, C. Y. "Extension of Multiple Range Tests to Group Means With Unequal Numbers of Replications," *Biometrics*, 12:307-310, 1956

Table A
MEAN SCORES, VARIANCES, AND STANDARD DEVIATIONS BY CLASSROOM UNITS FOR GRADES 12, 11, and 10

Grade 12

Regular					Trained					Movie					Audio				
N	$\overline{X}$	s^2	s	Sch.*	N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.
22	23.2	26.94	5.19	A	18	24.3	36.13	6.01	A	18	26.1	24.22	4.92	A	21	25.4	24.15	4.91	A
22	24.4	20.73	4.55	A	20	26.2	26.69	5.17	A	21	28.6	9.83	3.14	A	18	25.3	14.94	3.87	A
21	25.3	40.09	6.33	A	22	20.7	26.20	5.12	A	19	27.9	15.05	3.88	A	17	25.8	28.52	5.34	A
17	23.8	17.52	4.19	A	21	25.7	23.31	4.83	A	20	26.7	19.88	4.46	A	18	25.5	23.55	4.85	A
22	23.0	35.09	5.92	A	20	25.7	28.61	5.34	A	20	25.3	16.34	4.04	A	22	22.5	40.83	6.39	A

Grade 11

Regular					Trained					Movie					Audio				
N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.
21	23.4	34.85	5.90	A	32	22.5	25.16	5.02	A	20	25.5	34.47	5.87	A	20	23.8	21.32	4.62	A
22	25.5	30.64	5.54	A	18	21.8	34.33	5.86	A	22	23.1	30.44	5.52	A	22	24.3	18.52	4.30	A
21	24.9	33.54	5.79	A	21	25.0	21.69	4.66	A	22	23.9	29.22	5.41	A	20	23.1	28.30	5.32	A
21	25.7	22.81	4.78	A	21	22.9	21.44	4.63	A	20	25.1	23.77	4.87	A	21	24.8	16.32	4.04	A
22	22.6	31.94	5.65	A	23	23.0	32.36	5.69	A	22	24.5	32.83	5.73	A	21	26.1	20.16	4.49	A

Grade 10

Regular					Trained					Movie					Audio				
N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.
28	22.2	24.41	4.94	A	27	23.6	29.61	5.44	A	27	25.4	29.78	5.46	A	24	23.4	13.56	3.68	A
27	22.7	38.27	6.19	A	28	20.9	39.36	6.27	A	29	23.0	33.20	5.76	A	24	23.0	32.99	5.74	A
27	21.5	29.48	5.43	A	26	22.6	25.43	5.04	A	27	25.2	20.83	4.56	A	26	21.5	28.65	5.35	A
26	20.8	36.96	6.08	A	27	25.8	17.74	4.21	A	28	24.1	23.80	4.88	A	24	23.1	18.20	4.27	A
26	21.3	39.82	6.31	A	21	22.0	27.64	5.26	A						27	22.2	27.87	5.28	A

*School indicated by Letter (one high school)

Table B
MEAN SCORES, VARIANCES, AND STANDARD DEVIATIONS BY CLASSROOM UNITS FOR GRADES 9, 8, and 7

Grade 9

Regular					Trained					Movie					Audio				
N	$\overline{X}$	s^2	s	Sch.*	N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.
28	32.5	21.22	4.61	B	25	30.0	22.33	4.73	B	30	29.9	29.47	5.43	B	31	29.9	15.32	3.39	B
28	29.5	28.03	5.29	B	26	31.6	19.75	4.44	B	29	32.0	17.82	4.22	B	27	29.1	17.97	4.24	B
23	32.4	14.35	3.79	C	29	31.3	17.43	4.18	B	23	31.5	35.43	5.95	C	28	31.1	23.21	4.82	B
23	29.6	34.97	5.91	C	28	30.1	39.38	6.28	C	27	30.6	25.54	5.05	C	25	31.0	26.54	5.15	C
29	31.8	23.40	4.84	C	27	32.1	19.64	4.43	C	23	32.2	12.17	3.49	C	24	30.6	35.46	5.95	C

Grade 8

Regular					Trained					Movie					Audio				
N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.
28	30.4	29.73	5.45	B	26	29.6	30.22	5.50	B	25	32.6	29.82	5.46	B	28	28.8	27.58	5.25	B
28	29.7	30.73	5.54	B	28	28.4	27.66	5.26	B	27	29.4	33.40	5.78	B	25	29.2	28.96	5.38	B
30	29.6	21.69	4.66	B	28	31.1	25.95	5.09	C	25	29.5	25.75	5.08	B	26	30.7	17.70	4.21	C
27	31.7	20.73	4.55	C	26	30.8	30.13	5.49	C	23	31.1	18.95	4.35	C	24	29.5	30.26	5.50	C
25	28.6	25.24	5.02	C	28	28.4	24.48	4.95	C	27	29.3	19.61	4.43	C	27	29.5	40.26	6.35	C

Grade 7

Regular					Trained					Movie					Audio				
N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.	N	$\overline{X}$	s^2	s	Sch.
29	27.8	26.21	5.12	B	28	29.2	24.95	5.00	B	27	27.4	27.48	5.24	B	26	28.6	32.07	5.66	B
29	27.8	35.45	5.95	B	28	26.9	49.99	7.07	B	26	27.5	29.38	5.42	B	27	26.7	22.27	4.72	B
29	27.8	28.33	5.32	C	27	27.2	51.10	7.15	B	27	26.1	24.13	4.91	C	29	27.8	20.23	4.50	B
28	26.1	25.68	5.07	C	30	26.7	38.20	6.18	C	28	25.9	18.21	4.27	C	23	26.2	43.29	6.58	C
28	25.6	50.76	7.13	C	25	28.0	34.54	5.88	C	25	28.2	30.00	5.48	C	26	29.6	21.11	4.60	C

*Different Schools indicated by Letter (two junior high schools)

Table C
MEAN SCORES, VARIANCES, AND STANDARD DEVIATIONS BY CLASSROOM UNITS FOR GRADES 6, 5, and 4

Grade 6

Regular					Trained					Movie					Audio				
N	$\bar{X}$	s^2	s	Sch.*	N	$\bar{X}$	s^2	s	Sch.	N	$\bar{X}$	s^2	s	Sch.	N	$\bar{X}$	s^2	s	Sch.
23	34.4	17.62	4.20	E	23	35.5	7.62	2.76	E	29	32.4	17.18	4.15	D	23	32.5	14.89	3.86	H
25	33.9	9.37	3.06	J	25	35.2	6.85	2.62	E	27	31.6	28.62	5.35	D	19	30.8	20.80	4.56	H
29	33.4	14.89	3.86	J	28	32.5	26.03	5.10	G	22	33.9	14.71	3.84	F	23	31.6	36.52	6.04	I
26	36.4	4.73	2.18	K	27	34.5	4.95	2.23	G	22	34.5	10.16	3.19	F	19	33.1	16.98	4.12	I
25	34.3	17.47	4.18	K	22	33.9	8.04	2.84	H	27	34.6	9.31	3.05	K	25	33.9	7.91	2.81	K

Grade 5

Regular					Trained					Movie					Audio				
N	$\bar{X}$	s^2	s	Sch.	N	$\bar{X}$	s^2	s	Sch.	N	$\bar{X}$	s^2	s	Sch.	N	$\bar{X}$	s^2	s	Sch.
28	32.0	26.59	5.16	D	23	29.3	24.33	4.93	F	27	34.0	7.34	2.71	G	28	31.1	22.30	4.72	E
29	29.4	30.18	5.49	D	23	31.0	40.63	6.37	F	23	32.6	9.15	3.03	G	29	33.1	18.83	4.34	E
24	31.4	32.68	5.72	H	25	30.9	22.70	4.77	H	20	32.0	22.52	4.75	I	21	30.4	29.66	5.45	I
23	31.6	18.52	4.30	H	28	29.8	19.01	4.36	J	16	29.7	33.00	5.74	I	22	32.8	7.83	2.80	K
24	32.9	14.16	3.76	K	32	32.1	24.52	4.95	J	26	33.2	7.88	2.81	L	22	32.5	32.72	5.72	K

Grade 4

Regular					Trained					Movie					Audio				
N	$\bar{X}$	s^2	s	Sch.	N	$\bar{X}$	s^2	s	Sch.	N	$\bar{X}$	s^2	s	Sch.	N	$\bar{X}$	s^2	s	Sch.
30	29.5	21.70	4.66	G	28	28.5	39.88	6.32	D	28	32.9	13.84	3.72	E	27	28.6	27.84	5.28	F
31	30.6	25.22	5.02	G	30	27.3	43.06	6.56	D	25	31.7	30.44	5.52	E	29	26.7	31.38	5.60	F
30	28.8	28.97	5.38	I	28	29.8	24.15	4.91	E	27	28.7	31.33	5.60	H	26	26.9	41.35	6.43	H
27	31.5	12.09	3.48	K	22	29.1	25.45	5.05	I	27	29.7	21.73	4.66	H	24	31.0	20.34	4.51	J
28	31.5	21.06	4.59	K	20	28.5	34.89	5.91	I	28	32.1	22.57	4.75	K	22	30.6	16.51	4.06	J

*Different schools indicated by Letter (nine elementary schools).

Table D

MEAN TEST SCORES AND PERCENT OF CORRECT ANSWERS BY GRADES

High School — 36 Questions

Grade		Regular	Trained	Movie	Audio	Total	Publisher's Norms
12	Av. Correct	23.9	24.1	26.9	24.9	25.0	23.1
	Av. % Correct	66.4%	66.9%	74.7%	69.2%	69.4%	64.1%
11	Av. Correct	24.4	23.0	24.4	24.4	24.0	20.5
	Av. % Correct	67.8%	63.9%	67.8%	67.8%	66.9%	56.9%
10	Av. Correct	21.7	23.0	24.4	22.6	22.9	19.8
	Av. % Correct	60.3%	63.9%	67.8%	62.8%	63.6%	55.1%

Junior High School — 40 Questions

Grade		Regular	Trained	Movie	Audio	Total	Publisher's Norms
9	Av. Correct	31.2	31.0	31.2	30.3	30.9	27.8
	Av. % Correct	78.0%	77.5%	78.0%	75.8%	77.3%	69.6%
8	Av. Correct	30.0	29.7	30.4	29.5	29.9	26.3
	Av. % Correct	75.0%	74.3%	76.0%	73.8%	74.8%	65.7%
7	Av. Correct	27.4	27.6	27.0	27.8	27.5	23.9
	Av. % Correct	68.5%	69.0%	67.5%	69.5%	68.8%	59.7%

Elementary School — 40 Questions

Grade		Regular	Trained	Movie	Audio	Total	Publisher's Norms
6	Av. Correct	34.5	34.3	33.4	32.4	33.7	28.5
	Av. % Correct	86.3%	85.8%	83.5%	81.0%	84.3%	71.2%
5	Av. Correct	31.5	30.6	32.3	32.0	31.6	26.2
	Av. % Correct	78.8%	76.5%	80.8%	80.0%	79.0%	65.2%
4	Av. Correct	30.4	28.6	31.0	28.8	29.7	23.7
	Av. % Correct	76.0%	71.5%	77.5%	72.0%	74.3%	59.2%

Table E

VARIANCE RATIOS FOR ANALYSES OF VARIANCE WITHIN METHODS BY GRADE LEVELS*

Grade	Regular	Trained	Movie	Audio
12		3.55		
11				
10		3.28		
9				
8				
7				
6	2.80	3.32	2.81	
5			3.62	
4			3.47	3.66

*Variance ratios are given only if significant at the .05 level of confidence. Those which are underlined are also significant at the .01 level.

Table F

Individual Groups of Classes Wherein Significant Differences Were Found in Mean Scores Within Grade and Type of Administration

Grade	Type of Administration	Mean Score by Classes (Letter Represents School)				
12	TRAINED	A 20.7	A 23.7	A 24.3	A 25.7	A 26.2
10	TRAINED	A 20.9	A 22.0	A 22.6	A 23.6	A 25.8
6	REGULAR	J 33.4	J 33.9	K 34.3	E 34.4	K 36.4
6	TRAINED	G 32.5	H 33.9	G 34.5	E 35.2	E 35.0
6	MOVIE	D 31.6	D 32.4	F 33.9	F 34.5	K 34.6
5	MOVIE	I 29.7	I 32.0	G 32.6	L 33.2	G 34.0
4	MOVIE	H 28.7	H 29.7	E 31.7	K 32.1	E 32.9
4	AUDIO	F 26.7	H 26.9	F 28.6	J 30.6	J 31.0

Any two treatment means not underscored by the same line are significantly different (.05 level of confidence).

Any two treatment means underscored by the same line are not significantly different (.05 level of confidence).

Table G

Differences In Mean Scores Among The Methods of Administration By Test Level for High School and Junior High*

GRADES 10-12	REGULAR	TRAINED	AUDIO	MOVIE
$\bar{X}$ (Mean Score)	33.3	33.3	33.9	35.3

GRADES 7-9	AUDIO	TRAINED	REGULAR	MOVIE
$\bar{X}$ (Mean Score)	29.2	29.4	29.5	29.5

Any two treatment means not underscored by the same line are significantly different (.05 level of confidence).

Any two treatment means underscored by the same line are not significantly different (.05 level of confidence).

No analysis was made of mean scores for Grades 6-8 as a group because the interaction effect between grade level and method was significant at the .05 level.

*Based on Duncan's Multiple Range Test.

Table H

Differences in Mean Scores Among The Methods of Administration for Grades 6, 5, and 4*

GRADE 6	AUDIO	MOVIE	TRAINED	REGULAR
$\bar{X}$ (Mean Score)	32.4	32.4	34.2	34.5

GRADE 5	TRAINED	REGULAR	AUDIO	MOVIE
$\bar{X}$ (Mean Score)	30.6	31.5	32.0	32.2

GRADE 4	TRAINED	AUDIO	REGULAR	MOVIE
$\bar{X}$ (Mean Score)	28.6	28.8	30.4	31.0

Any two treatment means underscored by the same line are significantly different (.05 level of confidence).

Any two treatment means underscored by the same line are not significantly different (.05 level of confidence).

*Based on Duncan's Multiple Range Test.

Table I

RESULTS OF "t" TEST FOR DIFFERENCES BETWEEN CLASSES FROM THE SAME SCHOOL GIVEN THE TEST BY DIFFERENT METHODS

Grade	School	$\bar{X}$ (Mean) Regular	n	$\bar{X}$ (Mean) Trained	n	$\bar{X}$ (Mean) Movie	n	$\bar{X}$ (Mean) Audio	n	t
6	I	[illegible]	[illegible]	[illegible]	[illegible]					
6	II	[illegible]	[illegible]			[illegible]	[illegible]			[illegible]
6	III			[illegible]	[illegible]	[illegible]	[illegible]			[illegible]
6	IV					[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	V	[illegible]	[illegible]	[illegible]	[illegible]					[illegible]
6	VI	[illegible]	[illegible]					[illegible]	[illegible]	[illegible]
6	VII					[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	VIII	[illegible]	[illegible]	[illegible]	[illegible]					[illegible]
6	IX	[illegible]	[illegible]			[illegible]	[illegible]			[illegible]
6	X	[illegible]	[illegible]					[illegible]	[illegible]	[illegible]
6	XI							[illegible]	[illegible]	[illegible]
6	XII			[illegible]	[illegible]			[illegible]	[illegible]	[illegible]

*Significant at [illegible] level.

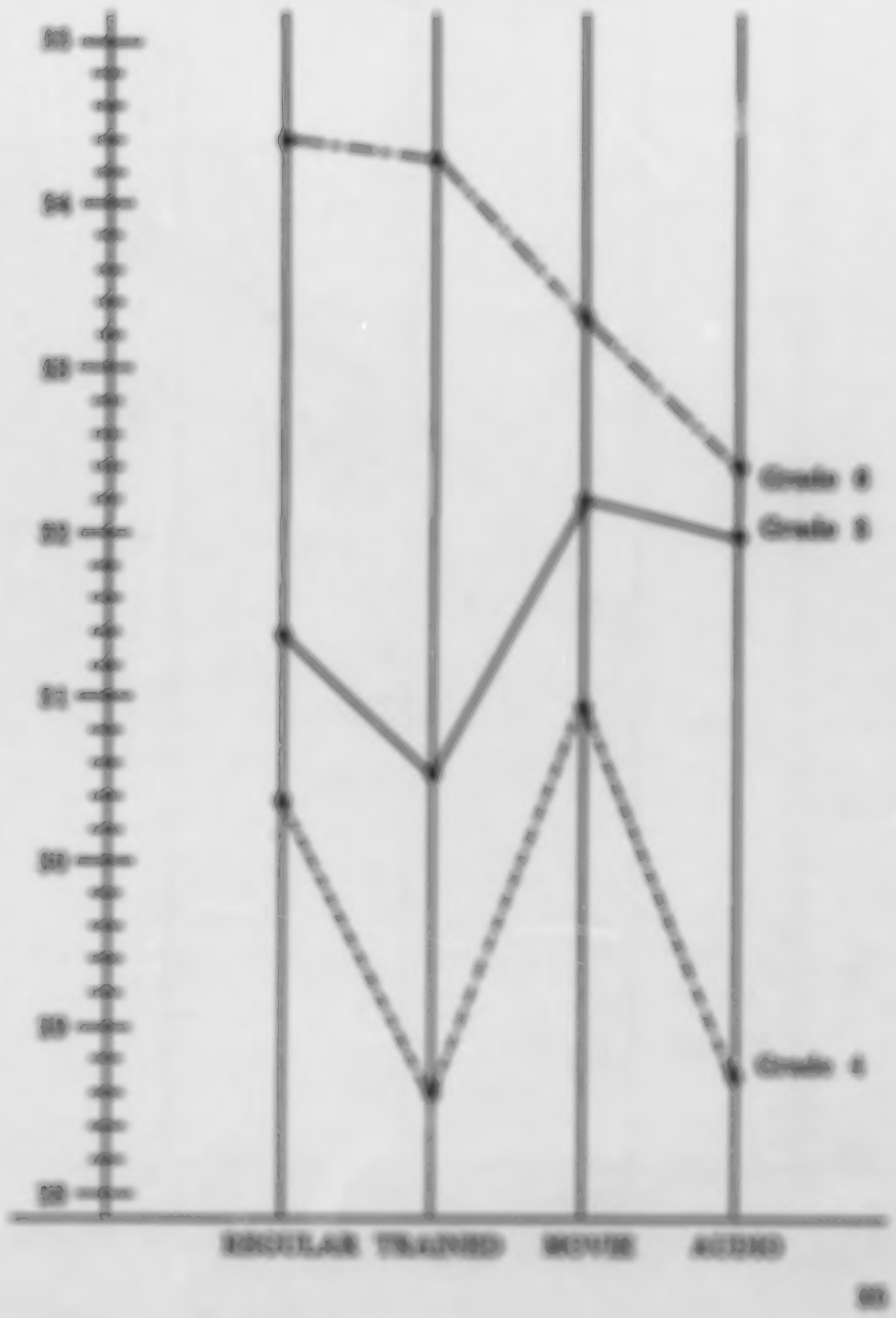
Table J

TYPES OF TAPE RECORDERS USED IN EXPERIMENT

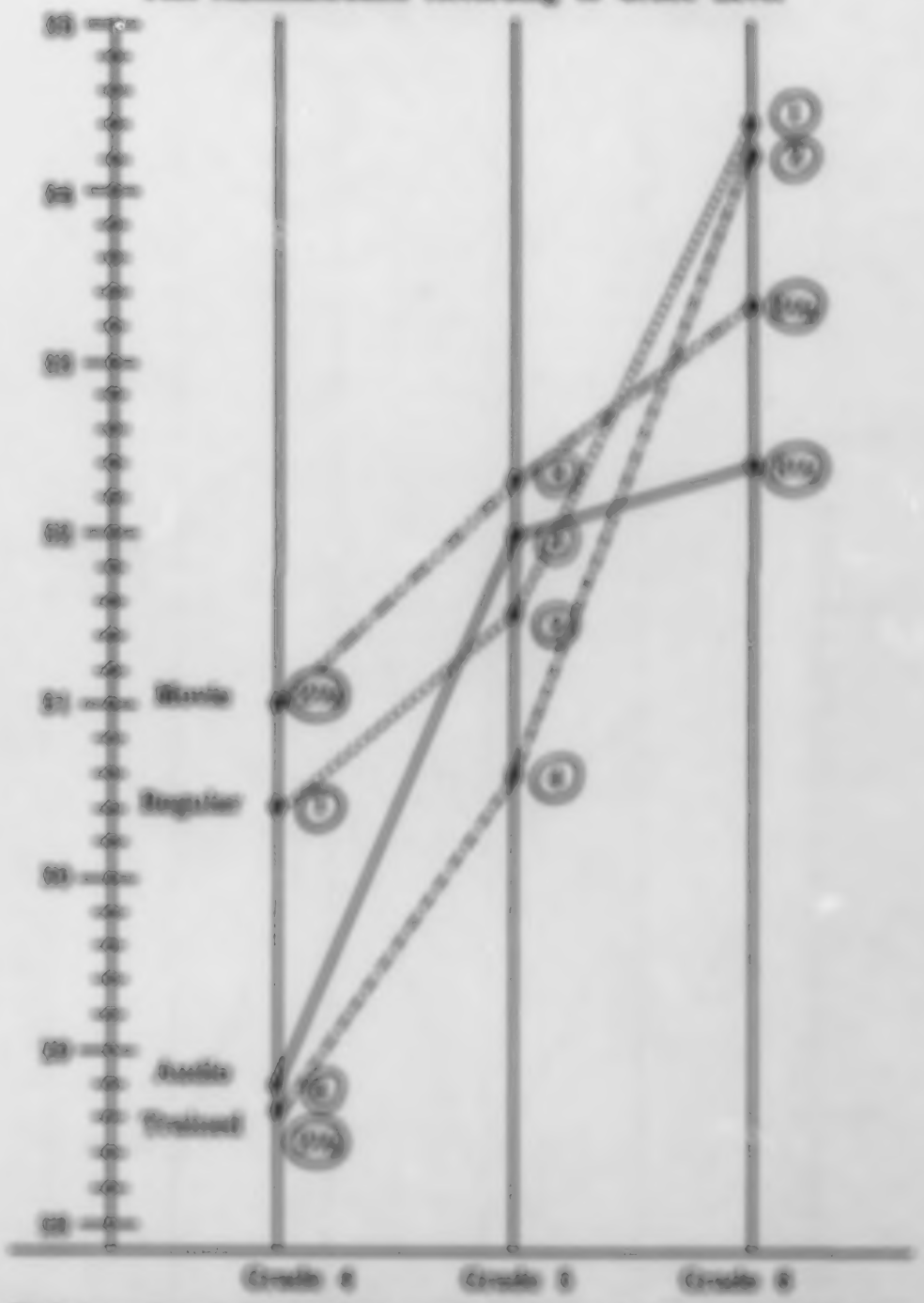
Type of Recorder	Number of Classes
Bell & Howell - 770-4	3
Bell & Howell - 73000	5
Continental	4
Magnecord	0
Magnecord Citation	4
Harvestone - A-130	1
Webcor - 110-1B	4
Webcor - 270-11	1
Webcor Royalette	0
Wollensak - T-1000	13
Wollensak - T-1000	4

Graph 1

Differences Among Scores for Grade Levels According to Methods of Test Administration



Graph 3
Differences Among Scores for Method of
Test Administration According to Grade Level



[illegible]

LIST OF TABLES AND GRAPHS

Page

Table I — [illegible]
Means and Variances of Listening Test Scores by Methods of Presentation for Grades 10, 11, and 12

Table II — [illegible]
Analysis of Variance by Methods of Administration and Grade Levels (10-12)
Method of Presentation for Grades 6, 8, and 9

Table III — [illegible]
Means and Variances of Listening Test Scores by

Table IV — [illegible]
Analysis of Variance by Methods of Administration and Grade Levels (7-9)

Table V — [illegible]
Means and Variances of Listening Test Scores by Method of Presentation for Grades 6, 8, and 9

Table VI — [illegible]
Analysis of Variance by Methods of Administration and Grade Levels (6-9)

Table VII — [illegible]
Mean Scores, Variances, and Standard Deviations by Classroom Units for Grades 10, 11, and 12

Table VIII — [illegible]
Mean Scores, Variances, and Standard Deviations by Classroom Units for Grades 6, 8, and 9

Table IX — [illegible]
Mean Scores, Variances, and Standard Deviations by Classroom Units for Grades 6, 8, and 9

Table X — [illegible]
Mean Test Scores and Percent of Correct Answers by Grades

Table XI — [illegible]
Variance Ratios for Analysis of Variance Within Methods by Grade Levels

Table XII — [illegible]
Individual Groups of Classes Whereby Significant Differences Were Found in Mean Scores Within Grade and Type of Administration

Table XIII — [illegible]
Differences in Mean Scores Among the Methods of Administration by Test Level for High School and Senior High

Table XIV — [illegible]
Differences in Mean Scores Among the Methods of Administration for Grades 6, 8, and 9

Table XV — [illegible]
Results of "T" Test for Differences Between Classes From the Same School Given the Test by Different Methods

Table XVI — [illegible]
Types of Type Extended Test in Experiment

Graph I — [illegible]
Differences Among Means for Grade Levels According to Methods of Test Administration

Graph II — [illegible]
Differences Among Means for Methods of Test Administration According to Grade Levels